SexualGA: Gender-Specific Selection for Genetic Algorithms

Stefan WAGNER
Institute for Formal Models and Verification
Johannes Kepler University
Linz, 4040, Austria

and

Michael AFFENZELLER
Institute for Formal Models and Verification
Johannes Kepler University
Linz, 4040, Austria

ABSTRACT

Selection for reproduction in the context of Genetic Algorithms uses only one selection scheme to select parent individuals. When considering the model of sexual selection in the area of population genetics it gets obvious that the process of choosing mating partners in natural populations is different for male and female individuals. In this paper the authors introduce a new selection paradigm for Genetic Algorithms (SexualGA) based upon the concepts of male vigor and female choice of population genetics which provides the possibility to use two different selection schemes simultaneously within one algorithm. By using this new concept it is possible to simulate sexual selection in natural populations more precisely. Furthermore, SexualGA also offers far more flexibility concerning the adaptivity of selection pressure enabling the GA user to tune the algorithm more accurately.

Keywords: Genetic Algorithms, Selection, Selection Pressure, Population Genetics

1. INTRODUCTION

Genetic Algorithms (GAs) developed by J. Holland in 1975 [14] represent a heuristic optimization technique inspired by the process of natural evolution. Like biological individuals different solution candidates form a population and develop from generation to generation by selection, crossover, and mutation. Thereby the essential steps can be stated as follows: The quality of each individual is evaluated by a fitness function leading to a so-called fitness value. This fitness value is the basis for the selection step in which above average fit individuals are chosen for reproduction. Crossover is then used to form new solution candidates out of the selected parents by combining the genetic code of two parents and in that way generating a new solution candidate called child in the terminology of GAs. Additionally, to avoid an early stagnation of the optimization algorithm mutation is used as a background operator to alter the gene material of the children a little bit with low priority (usually around 5%).

Although this basic concept of GAs is rather simple, GAs have proven to be a very powerful and flexible optimization strategy in numerous different applications like route planning, scheduling, time-table generation, or machine learning (cf. [12], [17], [7]). However, in comparison with nature, GAs are naturally still based on a drastic simplification of evolution. Moreover, GAs are also faced with some severe difficulties: Besides the problem of solution encoding and designing proper crossover and mutation operators, which is very critical to make the fundamental property of GAs hold (i.e. *"Good + Good = Even Better"*) [25], there is also a very sensitive interplay between forces reducing and maintaining genetic variety within the population (cf. [28], [21]). If selection or stochastic effects reduce genetic diversity too fast, the algorithm cannot procreate children outperforming their parents anymore. It reaches a state called premature convergence in GA theory which can be compared with the problem of getting stuck in a local but not global optimum well-known from neighborhood-based optimization techniques. Otherwise, if genetic diversity is not reduced sufficiently, the GA falls in some kind of equilibrium state being unable to exploit the full optimization potential stored in its population.

Different publications are available presenting various ways to better maintain genetic diversity (e.g. [6], [10], [12]), studying the phenomenon of premature convergence (e.g. [4], [3]), or analyzing the interplay of different forces inside GAs (e.g. [23]). However, the general concept of GA selection is rarely questioned. So in this paper the authors concentrate on this aspect. Thereby, the consideration of selection in natural populations and especially the area of population genetics gives valuable hints, how the classical GA selection model can be enhanced.

The area of population genetics concentrates on developing mathematical models to describe the variation of allele frequencies in natural populations (for a detailed overview see e.g. [13]). Especially different concepts of selection are thereby of importance, as selection manipulates the allele pool in a directed way. In that context sexual selection is described as the concept of male vigor and female choice, meaning that male individuals try to spread their gene material as wide as possible and female individuals are more selective by choosing rather above average fit males to guarantee a high survival probability of their offspring. So it seems to be preferable not to use identical selection mechanisms for male and female population members also in GAs.

Inspired by this view of sexual selection the authors present a new selection mechanism for GAs (SexualGA) by introducing two different selection operators, one for the selection of male and one for the selection of female individuals. In that way multiple combinations of different selection concepts can be realized, leading from strategies inspired by nature (like fitness proportional selection for choosing males and random selection for choosing females) to other combinations especially interesting in GA theory (like the combination of fitness proportional selection mechanisms with tournament or rank-based selection concepts).

In the experimental part of the paper the authors present results of various test runs with instances of the Traveling Salesman Problem (TSP) (e.g. [15]) taken from the TSPLIB [20]. These results highlight different effects of the new selection model concerning selection pressure and changing of genetic diversity and clearly show the higher flexibility provided by the SexualGA.

2. DIFFERENT SELECTION CONCEPTS IN POPULATION GENETICS

Population genetics aims to describe the topology and temporal dynamics of genetic variation in natural populations with the goal to understand the evolutionary forces that act on populations. Like Evolutionary Computation, population genetics also has an empirical as well as a theoretical component. Especially for scientists in the field of Evolutionary Computation, it should be a very fruitful approach to consider the latest developments of population genetics, which should be kept in mind as the bionic role-model for further developments. Therefore, in the following, some up to date considerations about population genetics is summarized which are relevant for the new selection concept described in this paper.

In the theory of GAs, selection and selection pressure are one of the most essential success criteria. But these factors support demographic stochasticity, i.e. the alleles of fitter individuals tend to dominate the population (as these individuals are selected for reproduction more often) whereas alleles which are stored only in individuals with lower fitness tend to get lost due to parent selection. Together with the fact that the size of the population is rather small in a typical GA, or in a certain deme of a distributed GA, this indicates the importance of thinking also about genetic drift as a genetic diversity reducing force [2] (cf. Hardy-Weinberg law of population genetics [13]). In population genetics the loss of genetic diversity due to selection and genetic drift is mainly considered adverse, as in nature genetic diversity in a population is essential for the ability to adapt to changing environmental conditions. However, from an Evolutionary Computation's point of view, the fixing of alleles containing genetic information of a global optimal solution is desirable. Vice versa also the loss of definitely disadvantageous alleles positively influences the performance of a GA. These considerations about the benefits of fixing or losing certain alleles in Evolutionary Algorithms is not respected in the diversity maintaining mechanisms discussed in GA theory [6], [10], [12].

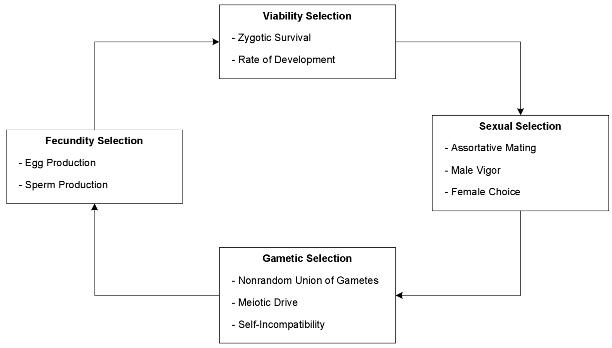

Figure 1: Various Aspects of Selection During the Lifecycle of an Individual

Since Darwin [8] selection is considered to be the most important (and only) evolutionary mechanism for adaptation to the environment. In natural populations selection occurs in a lot of different kinds like frequency-dependent selection, fecundity selection, gametic selection, viability selection, or sexual selection which are sometimes mathematically quite hard to handle. So in population genetics there are a lot of different selection models that try to deal with these aspects. The interplay between the different forms of selection during the lifecycle of an individual is visualized in Figure 1.

So obviously sexual selection which is selection for reproduction (i.e. the selection paradigm used in the context of GAs) is just one of many facets of selection in population genetics. Population genetics considers sexual selection mainly as interaction between male vigor and female choice, whereby the male individuals try to spread their genes in a rather broad and undirected way whereas the female individuals normally choose their partners depending on much harder criterions. So obviously there are gender-specific differences in the way mating partners are chosen in natural populations.

A very important consequence of selection in population genetics, as well as in evolutionary computation, is selection pressure and the resulting influence on certain alleles. Selection pressure occurs, as not all gene material contained in the parental population is also passed on and contained in the procreated children. Some of the genetic code gets lost. The reasons for that loss are manifold like the strict choice of mating partners, stochastic effects like genetic drift, stillbirths, or infant mortality. So as a matter of principle allele frequencies are not constant and steadily change from one generation to the next. In fact there are four possibilities for each allele in the population (where p denotes the relative frequency of an allele):

- $\mathbf{p} \rightarrow \mathbf{1}$: The allele is fixed in the entire population.
- $\mathbf{p} \rightarrow \mathbf{0}$: The allele is lost in the entire population.
- $\mathbf{p} \rightarrow \hat{\mathbf{p}}$: The allele frequency converges to an equilibrium state.
- $\mathbf{p} \rightarrow \mathbf{p}$: The allele frequency remains unchanged.

But when thinking of GAs and technical optimization problems the loss of genetic diversity and a fixing or losing of alleles is not considered as harmful as in natural population. Instead it is desirable that the gene pool of a GA converges to a very small allele set so that alleles which are part of a global optimal solution are fixed in the entire population and on the other hand alleles representing genetic information which is definitely not part of a global optimal solution are sieved out during the evolutionary optimization process.

3. SELECTION AND SELECTION PRESSURE IN GENETIC ALGORITHMS

In the context of GAs selection is only considered as sexual selection, also called selection for reproduction. The selection operator is used to choose two parents from the current generation in order to procreate a new child by crossover and mutation. In contrast to crossover and mutation, selection is problem independent per definition as only the fitness value of individuals is used. No information about the problem encoding or the problem itself is necessary.

Roulette Wheel Selection (RWS) is the most commonly used form of GA selection. When using RWS each individual is assigned a partition on a virtual roulette wheel which is equally large as its quality value. Then the wheel is spun and one individual is selected. Obviously, RWS is a fitness proportional selection scheme as the selection probability for each individual is directly proportional to its quality value. However, RWS has two major deficiencies: On the one hand it is very sensitive concerning super-individuals. A single individual with an unusual good fitness value (e.g. due to very successful mutation) is assigned a very long segment on the roulette wheel. Consequently its probability to be selected for reproduction is very high and the genetic material passed on to the next generation will be heavily dominated by this single individual (i.e. very high selection pressure). Such a strong domination causes a very high loss of genetic diversity which is definitively not advantageous for the optimization process. On the other hand Roulette Wheel Selection gets more and more undirected if the fitness values of the individuals become very similar. So especially in the later phases of the evolutionary search process it often drifts towards Random Selection.

As a consequence several other selection techniques with a probability not proportional to the individuals' fitness values have been developed. In general there are two different types of such non-proportional selection operators: First there are rank-based selections that assign the probability value depending on the order of the individuals according to their fitness values. The most common representative of this class is Linear Rank Selection (LRS) which is normally implemented in a lot-based way. The worst individual gets one lot, the next best two and so on up to the best individual getting as many lots as individuals are in the population. So selection probability is not influenced by super-individuals or the spreading of fitness values at all. Secondly, tournament-based selection techniques can be used where the best individual is selected out of a previously chosen group. In its most basic form a tournament group of n solutions is chosen randomly from the parental population which is called n-Tournament Selection (n-TS).

Besides their sensitivity concerning super-individuals and spreading of fitness values the selection schemes differ mostly in the selection pressure they cause. The concept of selection pressure was first introduced by Darwin [8] as a result of birth surplus: A population is producing more offspring than the actual environmental resources can keep alive. Consequently some of the not so fit children die before they reach the age of sexual maturity. This fact causes a so-called selection pressure among the offspring requiring a minimum fitness to survive in order to pass on their own genetic information. In the context of Evolutionary Computation selection pressure has been defined for some algorithms that also produce a birth surplus like Evolution Strategies (ES) [19], the Breeder GA [16], SEGA [1], or SASEGASA [4]. E.g. selection pressure is defined as $\frac{\lambda}{\mu}$ for the (μ, λ)-ES where μ denotes the population size and λ stands for the number of procreated offspring. A large value of $\frac{\lambda}{\mu}$ indicates a high selection pressure (small population size, lots of children) and vice versa. In the above mentioned algorithms selection pressure turned out to have a great influence on the algorithms' performance.

However, in the general case of GAs selection pressure cannot be defined so easily as a GA is procreating exactly as many children as needed. Nevertheless, numerous papers discuss selection pressure of GAs without measuring it in any way. E.g. it is said that Linear Rank

Selection or Tournament Selection cause a higher selection pressure than Roulette Wheel Selection. In fact there has to be something like selection pressure also in the case of classical GAs as it is not constantly easy or hard for an individual to pass on its own genetic information to the next generation. So the question is, how to make selection pressure measurable, if there is no birth surplus?

Selection pressure can be abstracted somehow as a measurement value indicating how hard it is for an individual to pass on its genetic characteristics from one generation to the next. So it seems reasonable to define selection pressure in a classical GA as the ratio between the population size and the number of individuals selected as parents for the next generation [23]. If the individuals of the next generation are procreated by a few parent individuals only, selection pressure is very high and vice versa. In the following empirical studies selection pressure is calculated according to Formula 1 where $|PAR|$ stands for the number of different selected parents and $|POP|$ represents the population size. So a minimum selection pressure of 0 indicates that all individuals of the parental generation got the chance of mating and a maximum selection pressure of $1 - \frac{1}{|POP|}$ represents the situation that all offspring are mutated clones of a single super-individual. However, in this definition of selection pressure it has to be considered that selection is still a stochastic process. Especially in small populations such stochastic effects occur quite often and can influence the algorithm's performance a lot. So the given formula is just an approximation. For an exact calculation of selection pressure infinitely large populations would have to be taken into account (cf. Hardy-Weinberg Law).

$$SP = 1 - \frac{|PAR|}{|POP|} \qquad (1)$$

So to analyze the differences of the commonly used selection schemes concerning selection pressure some experiments with the broadly established Standard Genetic Algorithm (SGA) (as described in e.g. [11], [9], [17], [18], [22]) attacking the `ch130` instance of the Traveling Salesman Problem (TSP) taken from TSPLIB [20] have been carried out by using the HeuristicLab optimization environment [24][1]. The used parameters are shown in Table 1. Table 2 lists the average selection pressure of the different selection schemes of ten independent runs. The performed experiments show, that selection pressure is totally invariant regarding the used crossover and mutation concepts. Furthermore, selection pressure is staying constant (except for stochastic effects) during the whole length of a GA run and is also not influenced by the size of the population even when using fitness proportional selection strategies. So obviously the level of selection pressure is really only dependent on the used selection scheme.

[1]In this paper only a small subset of the performed experiments can be presented. For a comprehensive overview see [23] or visit the HeuristicLab homepage `http://www.heuristiclab.com/results/afa`.

Table 1: Parameter Settings for the Test Runs

Generations	500
Population Size	150
Mutation Rate	0%
Crossover Operator	OX

Table 2: Average Selection Pressure for Different Selection Schemes

Selection Scheme	**Selection Pressure**	**Selection Scheme**	**Selection Pressure**
Random	0,13	6-TS	0,55
RWS	0,17	7-TS	0,59
LRS	0,24	8-TS	0,63
2-TS	0,24	9-TS	0,65
3-TS	0,36	10-TS	0,68
4-TS	0,44	20-TS	0,80
5-TS	0,50	50-TS	0,90

As a consequence the possibilities of a GA user to control selection pressure in a GA run are very limited. Only when using tournament selection, selection pressure can be affected by the used tournament group size rather coarsely. However, selection is the most relevant diversity reducing force in GAs and plays a very important role concerning the goal-directedness of the evolutionary search process. The fragile interplay between diversity supporting and diversity reducing forces is one of the most relevant aspects when thinking of the global solution quality achievable by a GA. So it would be very helpful for a GA user to have more control about the selection pressure caused by a selection scheme. These thoughts led to the development of the new selection concepts of the SexualGA which are described in the next section.

4. A NEW GENDER-SPECIFIC SELECTION PARADIGM

Inspired by the idea of male vigor and female choice as it is considered in the model of sexual selection discussed in the area of population genetics the authors developed a new selection paradigm for GAs called SexualGA. The main idea of the SexualGA is to use two different selection schemes for the selection of the two parents required for each crossover. So it gets possible to simulate the concept of male vigor and female choice by using random selection as the first selection scheme and another selection strategy with far more selection pressure as the second one (e.g. RWS or LRS). However, a separation of the GA population into a male and a female group cannot be done, as a diploid encoding would be necessary to differentiate sexes. Algorithm 1 shows the instruction

Table 3: Average Selection Pressure for SexualGA with Different Selection Schemes

	Random	RWS	LRS	2-TS	3-TS	4-TS	5-TS	6-TS	7-TS	8-TS	9-TS	10-TS
Random	**0,13**											
RWS	0,14	**0,17**										
LRS	0,15	0,19	**0,24**									
2-TS	0,16	0,20	0,24	**0,24**								
3-TS	0,18	0,21	0,29	0,29	**0,36**							
4-TS	0,20	0,22	0,32	0,32	0,39	**0,44**						
5-TS	0,22	0,23	0,33	0,34	0,42	0,47	**0,50**					
6-TS	0,23	0,24	0,35	0,35	0,43	0,49	0,53	**0,55**				
7-TS	0,24	0,25	0,36	0,36	0,44	0,50	0,54	0,57	**0,59**			
8-TS	0,25	0,26	0,36	0,36	0,45	0,51	0,55	0,58	0,61	**0,63**		
9-TS	0,26	0,27	0,37	0,37	0,46	0,52	0,56	0,59	0,62	0,65	**0,65**	
10-TS	0,27	0,27	0,37	0,37	0,46	0,53	0,57	0,60	0,63	0,66	0,67	**0,68**

sequence of SexualGA in pseudocode.

Algorithm 1 Sexual Genetic Algorithm (SexualGA)

```
Initialize total number of iterations nrOfIterations ∈ ℕ
Initialize size of population |POP|
Initialize mutation probability mutProb
Randomly produce an initial population POP_0 of size |POP|
Calculate fitness of each individual of POP_0
for i = 1 to nrOfIterations do
  Initialize next population POP_{i+1}
  while |POP_{i+1}| ≤ |POP| do
    Select father par_1 from POP_i by first selection scheme
    Select mother par_2 from POP_i by second selection scheme
    Generate a new child c from par_1 and par_2 by crossover
    Mutate c with probability mutProb
    Calculate fitness of c
    Insert c into POP_{i+1}
  end while
end for
```

So the SexualGA not only brings the concept of GAs a little bit more towards its biological archetype but also has relevant advantages compared to classical GA approaches particularly concerning flexibility. By using two different selection concepts simultaneously a GA user can influence the selection pressure level of a GA run more precisely. So it gets possible to control the interplay between genetic diversity supporting and reducing forces in a more directed way and to better tune GA behavior depending on the individual needs of the attacked optimization problem. To document the higher flexibility of SexualGA regarding selection pressure several test runs have been carried out using the same parameter set as already stated in Table 1 and also attacking the same problem instance. Again, the shown selection pressure values in Table 3 represent the average values of ten independent runs.

5. CONCLUSION

In this paper the authors introduced a new selection paradigm for GAs. Inspired by the concept of sexual selection in population genetics, where the process of choosing a mating partner is modeled in different ways for male and female individuals (male vigor and female choice), the SexualGA was developed. In contract to existing classical GA approaches, SexualGA uses two different selection schemes simultaneously for the selection of the two parents. So by combining Random Selection with selection strategies causing higher selection pressure the process of sexual selection in natural populations can be better simulated bringing GAs a little bit more towards their biological archetype.

In a comprehensive set of test runs the authors further showed that SexualGA also has severe advances concerning adaptivity of selection pressure. In the classical GA concept selection pressure can only be influenced by the size of the tournament group when using Tournament Selection. In contrast when using SexualGA selection pressure can be adjusted more precisely due to the combination possibilities of different selection concepts. So a GA user can better tune the algorithm regarding the needs of the attacked optimization problem. Furthermore, the optimization potential contained within a GA population can be exploited to a greater extent as the fragile interplay between diversity reducing and supporting forces can be adjusted more accurately.

Further research will focus primarily on the combination of SexualGA with already proposed advanced selection concepts for GAs (like Offspring Selection [5] and SASEGASA [4]). Especially when considering recent developments concerning structure identification and classification approaches based on Genetic Programming (see e.g. [26], [27]), the combination of Random Selection and fitness proportional selection schemes has already shown great potential.

6. REFERENCES

[1] M. Affenzeller. Segregative genetic algorithms (SEGA): A hybrid superstructure upwards compatible to genetic algorithms for retarding premature convergence. *International Journal of Computers, Systems and Signals (IJCSS)*, 2(1):18–32, 2001.

[2] M. Affenzeller. Population genetics and evolutionary computation: Theoretical and practical aspects. *Accepted to be published in: Journal of Systems Analysis Modelling Simulation*, 2004.

[3] M. Affenzeller and S. Wagner. SASEGASA: An evolutionary algorithm for retarding premature convergence by self-adaptive selection pressure steering. In J. Mira and J. R. Alvarez, editors, *Computational Methods in Neural Modeling*, volume 2686 of *Lecture Notes in Computer Science*, pages 438–445. Springer-Verlag, 2003.

[4] M. Affenzeller and S. Wagner. SASEGASA: A new generic parallel evolutionary algorithm for achieving highest quality results. *Journal of Heuristics - Special Issue on New Advances on Parallel Meta-Heuristics for Complex Problems*, 10:239–263, 2004.

[5] M. Affenzeller and S. Wagner. Offspring selection: A new self-adaptive selection scheme for genetic algorithms. In *Adaptive and Natural Computing Algorithms*, pages 218–221. Springer, 2005.

[6] D. Cavicchio. *Adaptive Search Using Simulated Evolution.* PhD thesis, University of Michigan, 1975.

[7] L. Chambers, editor. *Practical Handbook of Genetic Algorithms*, volume 2. CRC Press, 1995.

[8] C. Darwin. *The Origin of Species.* Wordsworth Classics of World Literature. Wordsworth Editions Limited, 1998.

[9] L. Davis. *Handbook of Genetic Algorithms.* Van Nostrand Reinhold, 1991.

[10] K. A. DeJong. *An Analysis of the Behavior of a Class of Genetic Adaptive Systems.* PhD thesis, Department of Computer Science, University of Michigan, 1975.

[11] D. Dumitrescu, B. Lazzerini, L. C. Jain, and A. Dumitrescu. *Evolutionary Computation.* The CRC Press International Series on Computational Intelligence. CRC Press, 2000.

[12] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison Wesley Longman, 1989.

[13] D. L. Hartl and A. G. Clark. *Principles of Population Genetics.* Sinauer Associates Inc., 2^{nd} edition, 1989.

[14] J. H. Holland. *Adaption in Natural and Artifical Systems.* University of Michigan Press, 1975.

[15] E. L. Lawler, J. K. Lenstra, A. Rinnooy-Kan, and D. B. Shmoys. *The Travelling Salesman Problem.* Wiley, 1985.

[16] H. Mühlenbein and D. Schlierkamp-Voosen. The science of breeding and its application to the breeder genetic algorithm BGA. *Evolutionary Computation*, 1(4):335–360, 1993.

[17] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs.* Springer-Verlag, 1992.

[18] M. Mitchell. *An Introduction to Genetic Algorithms.* The MIT Press, 1996.

[19] I. Rechenberg. *Evolutionsstrategie.* Friedrich Frommann Verlag, 1973.

[20] G. Reinelt. TSPLIB - A traveling salesman problem library. *ORSA Journal on Computing*, 3:376–384, 1991.

[21] R. E. Smith, S. Forrest, and A. S. Perelson. Population diversity in an immune systems model: Implications for genetic search. In *Foundations of Genetic Algorithms*, volume 2, pages 153–166. Morgan Kaufmann Publishers, 1993.

[22] M. Tomassini. A survey of genetic algorithms. *Annual Reviews of Computational Physics*, 3:87–118, 1995.

[23] S. Wagner. *Looking Inside Genetic Algorithms.* Schriften der Johannes Kepler Universität Linz, Reihe C: Technik und Naturwissenschaften. Universitätsverlag Rudolf Trauner, 2004.

[24] S. Wagner and M. Affenzeller. Heuristiclab: A generic and extensible optimization environment. In *Adaptive and Natural Computing Algorithms*, pages 538–541. Springer, 2005.

[25] S. Wagner, M. Affenzeller, and D. Schragl. Traps and dangers when modelling problems for genetic algorithms. In R. Trappl, editor, *Cybernetics and Systems 2004*, volume 1, pages 79–84. Austrian Society for Cybernetic Studies, 2004.

[26] S. Winkler, M. Affenzeller, and S. Wagner. Identifying nonlinear model structures using genetic programming techniques. In R. Trappl, editor, *Cybernetics and Systems 2004*, volume 1, pages 689–694. Austrian Society for Cybernetic Studies, 2004.

[27] S. Winkler, M. Affenzeller, and S. Wagner. New methods for the identification of nonlinear model structures based upon genetic programming techniques. In Z. Bubnicki and A. Grzech, editors, *Proceedings of the 15^{th} International Conference on Systems Science*, volume 1, pages 386–393. Oficyna Wydawnicza Politechniki Wroclawskiej, 2004.

[28] Y. Yoshida and N. Adachi. A diploid genetic algorithm for preserving population diversity - pseudomeiosis ga. In *Lecture Notes in Computer Science*, volume 866, pages 36–45. Springer-Verlag, 1994.